The Misuse of Statistics: Concepts, Tools, and a Research Agenda

John S. Gardenier, National Center for Health Statistics, Centers for Disease Control and Prevention, USA

David B. Resnik, Brody School of Medicine, East Carolina University, USA

Key Words. Competence, Ethics, Journals, Research misconduct, Statistics

While it is widely recognized that the proper use of statistics is a key element of research integrity, there has been considerable debate about how to understand or respond to the *misuse* of statistics in research. To understand what is meant by "misusing statistics," it is important to describe the role of statistics in the scientific method and relate the concept of "misuse" to other ethical concepts, such as "misconduct" or "incompetence" or "negligence." We believe that some misuses of statistics can be considered misconduct, although most misuses should be viewed as negligence or deficits of competence.

Statistical methods, theory, techniques, and models play an important role in several stages of the scientific method, but we will focus here on just two stages (See Figure 1). First, statistics is essential to good *experimental design* as in randomized clinical trials, for example. In order to obtain a rigorous test of a hypothesis, it important to obtain data that can provide evidence for or against the hypothesis. If the hypothesis is a comparative or quantitative statement, such as "drug x is more effective than drug y" or "less than five percent of patients suffer serious side effects from drug x," then the conclusions must be based on statistically significant results. For example, an experiment that compares the effects of two drugs on only ten patients is very unlikely to produce statistically significant results. If some or all of those patients are subjected to health risks in the experiment, this creates two additional ethical problems. First, it is unethical to expose a human subject to an unnecessary experimental risk, unless the potential benefits (to the individual or to society) of exposure to the risk outweigh the potential harms. If the experiment is not well designed such that no meaningful conclusions can be drawn, then the potential benefits will not outweigh the potential harms. Second, when patients give informed consent to participate in research, they usually believe that the research is valuable and may advance science. Encouraging or even allowing subjects to participate in an experiment that is highly unlikely to yield valid results is implicitly deceptive. It is important to address the statistical issues before conducting experiments or tests, because once one has gathered and recorded data, it may be too late to correct statistical (or ethical) flaws in the design of the experiment (1). The expression "garbage in, garbage out" applies here.

Second, statistics is important in an *analyzing and interpreting data*. There are many different statistical tools that one may use to analyze data, ranging from simple procedures, such as t-tests and

---

Corresponding author: John S. Gardenier, CDC/NCHS, 6525 Belcrest Road, Hyattsville, MD 20782, 301-458-4229 (voice), 301-458-4021 (fax), jsg6@cdc.gov.

* Disclaimer: This paper represents the personal views of the authors. It does not necessarily represent policies or positions of their institutions. The research effort reported here was not supported by grant funds.

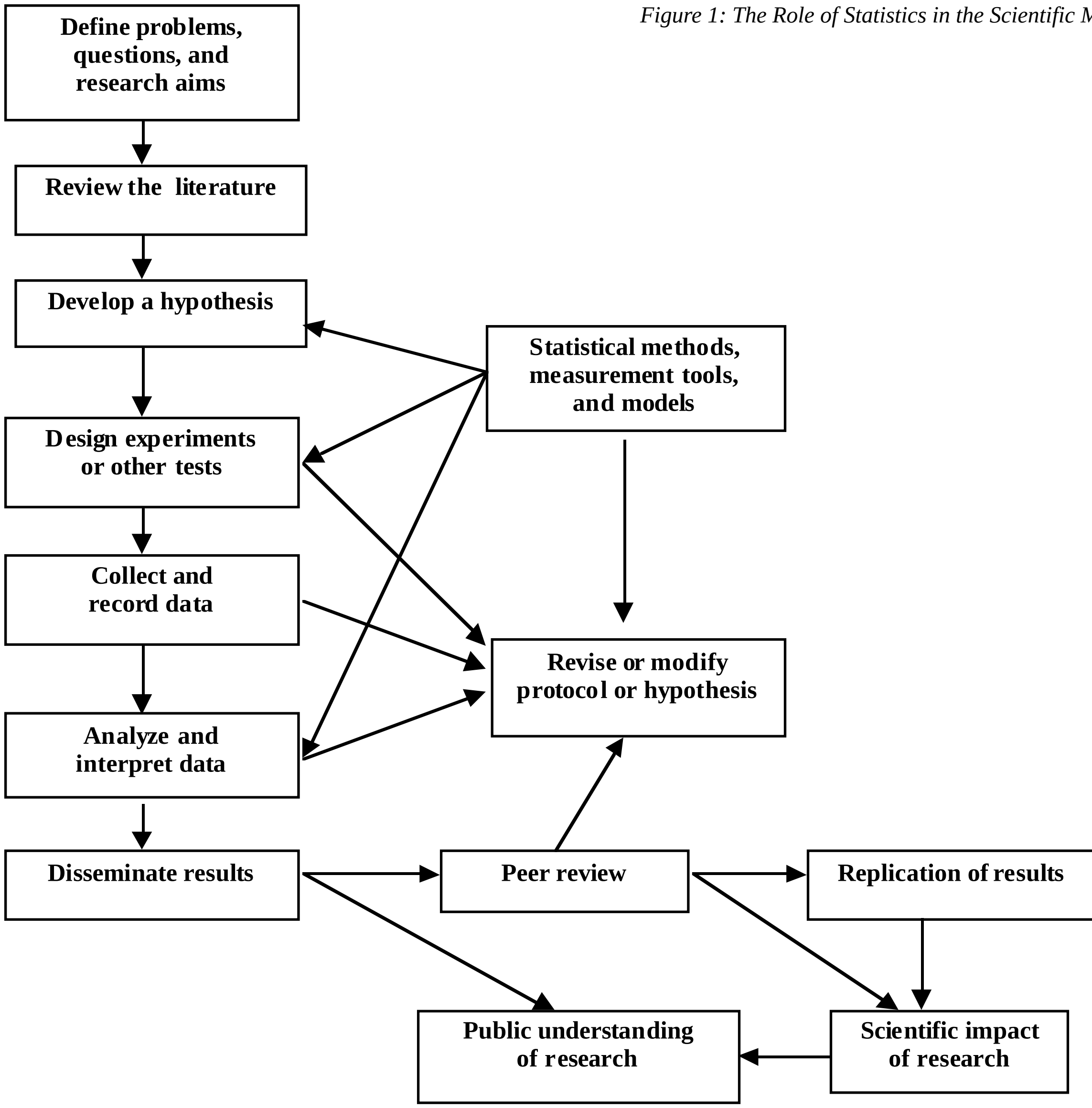

Figure 1: The Role of Statistics in the Scientific Method

linear regression, to more complex ones, such as analysis of covariance and statistical modeling. It is not our aim to discuss these methods here, but we would like to point out that it is relatively easy to misuse these methods. To apply any statistical method correctly, one must have information about the variables used (continuous or discrete, gaussian or bimodal, etc.), information about the sampling process used (sample size, independence, randomness, representativeness, etc.), and a sound understanding of the theory and assumptions underlying that method. If a researcher does not use a method correctly, then conclusions may overestimate or underestimate an important relationship or effect. If we think of statistics as a tool for distinguishing between random "noise" in the data and the real signal, then someone who incorrectly uses statistics may produce a result that is distorted or even artificial. A person who correctly uses statistics will amplify and clarify the signal without distorting it (2).

With this understanding of the role of statistics in research in mind, we can clarify what we mean by "misuse" of statistics. Not all misuses have equivalent ethical implications, as we discuss later. A "misuse," for our purposes, is an *incorrect* use, i.e., a use of statistics that is not appropriate, given the research question, the experimental design, and the methods being used. For example, it may be appropriate to exclude outliers if there is credible evidence that such points are not part of the statistical population represented by the sample. It may also be appropriate to use statistical methods to fill in (or impute) missing data for the purposes of statistical analysis. What's the difference between appropriate and inappropriate exclusion of outliers or appropriate and inappropriate imputation of data? Many books on statistical methods discuss these topics, but from an ethical viewpoint they boil down to the following: an appropriate exclusion (or imputation) is one that dampens the noise without altering the signal that describes the relationship or effect.

Misuses of statistics can also occur in the absence of erroneous or distorted results. Misuse can also arise from a failure to provide the research community with *important information* about the methods used or the experimental design. Researchers need to address such statistical issues as excluding outliers, imputing data, editing data, "cleaning" data, or "mining data."[2] These practices are often practical, or even necessary, but it is important to discuss them honestly and openly when reporting research results (3).

Thus, there are two types of misuses in statistics: (1) using statistical methods, techniques, or models in ways that produce distorted or artificial results; (2) failing to disclose important information about statistical methodology to researchers. Misuses of statistics may (or may not) violate several ethical obligations, such as the duty to be honest, the duty to be objective, the duty to avoid error, and possibly the duty to be open (4). There has been considerable debate about whether "misuse of statistical methods" should be classified as misconduct (5). The federal government and the scientific community have moved toward a narrow definition of misconduct that focuses on *fabrication, falsification, and plagiarism* (6, 7). The new federal policy implies that the misuse of statistics could be classified as a form of misconduct when it involves *intentional deception*. Some misuses could be classified as "fabrication" if they involve making up data or results, or "falsification" if they involve manipulating, changing, or omitting data or results. Misuses of statistics that do not involve intentional deception could be viewed as honest error, incompetence, bias, or "serious deviations" from acceptable practice (8). A person who makes excessive errors due to haste, ignorance, or sloppiness may be considered to be negligent or lacking the needed degree of competence, statistical or otherwise (9). Professionalism requires adequate application of both statistical and subject matter expertise to analyses. There might be varying degrees of culpability in a failure to meet this criterion. Clearly, honest error is never misconduct. Neither is it misconduct when two or more well qualified statisticians or other researchers disagree about technical issues in a given research protocol. Still, some misuses of statistics in research do fit the definition of misconduct used by the federal government. That may be hard to establish by a preponderance of the evidence. When a person changes or fabricates data, one at least has some kind of record that one can use to imply intent. When a person manipulates analyses of data, there may be no record to prove the manipulation was deliberate or even culpably negligent. Thus, as a purely practical matter, it may be very difficult investigate or prosecute such cases (10).

The Importance of Correcting Statistical Misuse

Statistics play vital roles in most aspects of modern post-industrial societies. Although statistics are sometimes dismissed as trivia or fuzzy math, distrusted as biased, or directly equated with lying, the truth is that they are inescapably important (11). As noted in the Preamble to the Ethical Guidelines for Statistical Practice:

> The professional performance of statistical analyses is essential to many aspects of society. The use of statistics in medical diagnoses and biomedical research may affect whether individuals live or die, whether their health is protected or jeopardized, and whether medical science advances or gets sidetracked. Life, death, and health, as well as efficiency, may be at stake in statistical analyses of occupational, environmental, or transportation safety. Early detection and control of new or recurrent infectious diseases depend on sound epidemiological statistics. Mental and social health may be at stake in psychological and sociological applications of statistical analysis.
>
> Effective functioning of the economy depends on the availability of reliable, timely, and properly interpreted economic data. The profitability of individual firms depends in part on their quality control and their market research, both of which should rely on statistical methods. Agricultural productivity benefits greatly from statistically sound applications to research and output reporting. Governmental policy decisions regarding public health, criminal justice, social equity,

> education, the environment, the siting of critical facilities, and other matters depend in part on sound statistics.
>
> Scientific and engineering research in all disciplines requires the careful design and analysis of experiments and observations. To the extent that uncertainty and measurement error are involved – as they are in most research – research design, data quality management, analysis, and interpretation are all crucially dependent on statistical concepts and methods. Even in theory, much of science and engineering involves natural variability. Variability, whether great or small, must be carefully examined both for random error and for possible researcher bias or wishful thinking. . . .
>
> Because society depends on sound statistical practice, all practitioners of statistics, whatever their training and occupation, have social obligations to perform their work in a professional, competent, and ethical manner. (12)

If researchers are careless or deceptive in their use of statistics, harms and costs to society will result. Poor statistics in science leads to poor science. The research record can be corrupted or polluted, wasting the time and energy of other researchers. At the very least, research funds lost in bad research represent an opportunity cost in that those funds could have been allocated to more deserving projects.

For all of these reasons, it is important that scientists and science administrators pay careful attention to the quality of statistics in science as funded, performed, and reported in their areas of jurisdiction and of competence. Good statistical work should be defended when it is attacked inappropriately. Bad statistical work should be detected and corrected as appropriate.

What are the Contributing Factors to Misuse?

There is not a great deal of evidence that has a direct bearing on the misuse of statistics in research. However, if one assumes that many of the factors that contribute to other ethical problems in research, such as misconduct, probably also play a role in the misuse of statistics, then one could cite the following factors, i.e., the "usual suspects" (13, 14).

- Pressures to publish, produce results, or obtain grants
- Career ambitions or aspirations
- Conflicts of interest and economic motives
- Inadequate supervision, education, or training

We believe that all of these factors probably play a role in misuses of statistics, but our conclusions are merely speculative. More research is needed on this topic. However, we would like to discuss two other possible factors in the misuse of statistics that are not on the above list of "usual suspects."

First, there are now many computer programs that analyze data. These programs are very user-friendly; all you need to do is load your data set and choose your statistical test in order to get results. One may even run several different tests in an attempt to increase the significance level (or p-value), although this can invalidate the testing. While these programs save a great deal of time and effort, they may contribute to statistical misuse in that it is possible to plug some numbers into one of these programs without knowing how the analysis works, or why a certain test would (or would not) be an appropriate test. We think this problem has a fairly obvious solution: teach more statistics in research. If students and researchers understand how to use statistics properly, then they should have fewer problems using statistical computer programs. Indeed, we believe that education is the key to improving statistical practice.

Second, it has become standard practice in some areas of research to only publish results that have a p-value of 0.05 or less. The best journals use more comprehensive criteria enforced by competent statistical peer review. We here address only those journals that place excessive reliance on the p-value. The value of 0.05 is an arbitrarily chosen number; there is no sound statistical or philosophical reason why a p-value of 0.06 is fundamentally different from a p-value of 0.05. However, under pressure to publish, researchers may decide to massage or manipulate data in order to obtain "significant" results. Furthermore, there is now a growing body of literature on publication bias in research (15-17). Publication bias occurs when there are discrepancies between the published research record and the complete research record. The discrepancies occur because journals tend to publish only "significant" results. There are some good potential solutions to the p-value problem. First, researchers should realize that p-values are merely conventional, not sacrosanct.

Second, they are also often sensitive to various theoretical assumptions and may give erroneous results due to mere artifacts of a data sample. Third, not all statistical computer packages compute all tests correctly. Fourth, journals should be willing to publish results that are substantial contributions to the literature of the field, not just those that appear to have met a conventional statistical test. The test result reported may not be correct, and even a correct conclusion that a certain hypothesis was not statistically supported by data from a well-designed study may be useful in limiting future fruitless research by others. Finally, researchers and research organizations should create databases for unpublished data of archival value and make those data publicly available (18).

Statistical Ethics, a Powerful Tool for Research Integrity

Statistical ethics is a relatively recent development. The seminal work, by W. Edwards Deming, was first published in 1965 (19). The American Statistical Association developed a series of statistical ethics codes or guidelines starting in 1979. Their current official Ethical Guidelines for Statistical Practice was promulgated in 1999 (12). The International Statistical Institute instituted its Declaration on Professional Ethics in 1985 (20). The United Nations has published Fundamental Principles of Official Statistics in the early 1990s, the current official version being dated 1994 (21).

The pattern that emerges from this brief history is that initial efforts to approach the issue tend to be optimistically simple. Corrections over time add to the scope and complexity of the documents. The most recent document breaks out areas of ethical responsibility for all people using statistical methods professionally (12). It covers separately, for example, responsibilities in publications and testimony, responsibilities to funders or employers, to research subjects, to research team colleagues, and responsibilities regarding allegations of misconduct. Beyond addressing responsibilities of the individuals, moreover; it also addresses the responsibility of those employing practitioners of statistical methods to provide a suitable moral climate for that work.

Such statistical ethics documents become tools for research integrity when they are integral to actual practice. For example, if a federal research funding agency were to adopt a policy of stating in grant announcements that all grant proposals received for projects employing statistical methods would be expected to be performed in accordance with the Ethical Guidelines for Statistical Practice, that would put real moral pressure on both proposers and grantees to avoid misuse of statistics. If journal editors were to state in notices to authors that any papers containing statistical methods submitted to that journal would be implicitly subject to those same guidelines, some of the authors would be more highly motivated to avoid misuse of statistics.

If all scientists and engineers who are competent in statistical methods would note published examples of misuse of statistics and report those to the funding agencies or journal editors involved, then the recipients would become more motivated to enforce sound statistical practice. In short, we should not let ethics documents sit unused on shelves or in unvisited cyberspace. Ethical considerations have practical consequences for good or evil. The failure of good people to use them effectively contributes to the likelihood that other people may perpetuate statistical misuse either through intent to deceive or simply through deficits of statistical competence.

A Proposed Research Agenda

While we believe that there are still many important conceptual and theoretical issues relating to the use/misuse of statistics in research, it should be clear from this brief discussion that more empirical research is required on the incidence of statistical misuse, its causes and effects¸ and on the efficacy of using ethics education and ethics documents as tools for improvement. The following are some of the empirical research questions we think are important to study:

1. How many (or what percentage of) published studies make statistical mistakes?
2. How many allegations of research misconduct involve misuses of statistics?
3. How many researchers believe that the misuse of statistics is an important ethical issue in research?
4. Do different fields have different statistical practices or take different approaches to the misuse of statistics?

5. What is the incidence of publication bias in various fields?
6. What do researchers and students know about statistics?
7. Where, when, and how do students learn about misuses of statistics in research or other ethical issues in statistics?
8. How often do researchers use statisticians or other statistical consultants?
9. Are editors and reviewers able to catch statistical misuses?
10. Can data audits detect misuses of statistics?
11. Do research ethics codes or policies address misuses of statistics?
12. When ethics education or ethics documents are used as tools to improve research integrity, how effective are they at promoting the proper use of statistics?
13. How often do institutional review boards (IRBs) discuss statistical issues in human subjects research? Do IRBs use statisticians?
14. How do misuses of statistics affect the public? Do such misuses ever cause harm to the public or threaten public health or safety?
15. How often do statistical issues arise in public policy debates?
16. What does the public know (or not know) about statistics?
17. How do lay people interpret important statistical concepts, such as "probability," and "risk"?

Bibliography

1. Johnson R. and Bhattacharyya G. *Statistics: Principles and Methods*. New York: John Wiley and Sons, 1985.
2. DeMets D. Statistics and ethics in medical research. *Science and Engineering Ethics* 1999*;* 5: 97-117.
3. Resnik D. Statistics, ethics, and research: an agenda for education and reform. *Accountability in Research* 2000; 8: 163-88.
4. Bailar J. Science, statistics, deception. *Annals of Internal Medicine* 1986; 104: 259-60.
5. Guston D. Changing explanatory frameworks in the U.S. government's Attempt to define research misconduct. *Science and Engineering Ethics* 1999; 5: 137-54.
6. Office of Research Integrity. Federal policy on research misconduct. Federal Register: October 14, 1999; 64, 198: 55722-55725.
7. Office of Science and Technology Policy. Federal policy on research misconduct. *http://www.ostp.gov/html/001207_3.html* , 2000.
8. Panel on Scientific Responsibility and the Conduct of Research. *Responsible Science: Ensuring the Integrity of the Research Process*. Washington: National Academy Press, 1992.
9. Grinnell F. *The Scientific Attitude*. New York: The Guilford Press, 1992.
10. Dahlberg, J. Personal communication. Office of Research Integrity, Rockville, MD. 20 November 2000.
11. Huff D. *How to Lie with Statistics*. New York: WW Norton, 1954.
12. American Statistical Association. *Ethical Guidelines for Statistical Practice*. Alexandria, VA: American Statistical Association, 1999. Available at *http://amstat.org/profession/ethicalstatistics.html.*
13. Macrina, F. *Scientific Integrity*. Washington: American Society for Microbiology Press, 2000.
14. Commission on Research Integrity. *Integrity and Misconduct in Research*. Washington: Public Health Service, 1995.
15. Ioannidis J. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998*;* 79: 281-86.
16. Stern J and Simes R. 1997. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; 315: 640-45.
17. Easterbrook P, et al. Publication bias in clinical research. *Lancet* 1991; 337: 867-72
18. Rennie D. Fair conduct and fair reporting of clinical trials. *JAMA* 1999; 282: 1766-68.
19. W. Edwards Deming. Principles of Professional Statistical Practice. *The Annals of Mathematical Statistics* 1965; 36: 1883-1900.
20. International Statistical Institute. Declaration on Professional Ethics. 1985. Available at *http://www.cbs.nl/isi/ethics.htm*.
21. United Nations. Fundamental Principles of Official Statistics. 1994. Available at *http://www.cbs.nl/isi/fundamental.htm* .